# Software Measurement Frameworks to Assess the Value of Independent Verification & Validation

**Dr. Nancy Eickelmann**
**NASA IV&V Facility**
**Software Research Laboratory**
**100 University Drive**
**Fairmont, West Virginia 26554**
**+1 304 367 8444**
**http://research.ivv.nasa.gov/~ike**
Nancy.Eickelmann@ivv.nasa.gov

**Abstract**

Software IV&V, as practiced by the NASA IV&V Facility, is a well-defined, proven, systems engineering discipline designed to reduce risk in major software systems development. However, we currently have no proven methodology for estimating resource requirements for IV&V based on sound financial criteria. The quantification of a cost structure associated with IV&V and the resulting benefits are essential to make objective decisions concerning the allocation of resources to IV&V activities. The development of ROI metrics for NASA IV&V would provide key information to make rational budgetary decisions that impact safety and mission critical aspects of all NASA software systems. To measure IV&V benefits and costs we must identify relevant measures and provide target ranges for those measures that may be used to evaluate whether or not the goals are achieved and to what degree. This requires a measurement strategy for software IV&V in the NASA context. This paper presents the NASA IV&V Balanced Scorecard strategic measurement framework and discusses its role in providing a minimal and usable core metrics set.

## 1 Introduction

The Balanced Scorecard, as applied in industry and government, is approached from two very disparate viewpoints. Industry is very aware of the importance of financial performance measures in managing an organization. Publicly held companies must be responsive to market and shareholder demands. Market share, share price, dividend growth, and other significant results-oriented financial measures have been used historically to evaluate an organization. Government organizations must respond to regulatory and legislative acts. One such legislative act is the Government Performance and Results Act (GPRA) passed by Congress and signed by the President in 1993. This act provides a new tool to improve the efficiency of all Federal agencies.

The goals of GPRA are to:

- Improve Federal program management, effectiveness, and public accountability
- Improve congressional decision making on where to commit the Nation's financial and human resources
- Improve citizen confidence in government performance

A specific difference between government and industry is explicit in the government's focus on cost reduction as compared to industry's focus on revenue generation and profitability. We have customized our BSC to accommodate these differences thus providing a framework to evaluate the overall performance of the organization through a linked hierarchy of specific performance drivers and outcome measures [7].

### 1.1 Structure of the Paper

Section 2 provides an overview of the Balanced Scorecard and motivations for its use. We then excerpt portions of our scorecard to exemplify our measurement framework, the application of cause effect graphing and the setting of strategic measurement targets in Section 3. Section 4 discusses specific BSC measurement issues and lesson learned. Section 5 concludes our paper and discusses current directions of our work.

## 2 Balanced Scorecard

The Balanced Scorecard (BSC) Framework provides the necessary structure to evaluate quantitative and qualitative information with respect to the organization's strategic vision and goals. There are two categories of measures used in the BSC the leading indicators or performance drivers and the lagging indicators or outcome measures. The performance drivers or leading indicators enable the organization to quantitatively track whether or not the organization is achieving short-term operational improvements. The outcome measures or lagging indicators provide objective evidence of whether *strategic objectives* are achieved and to what degree. The two measures must be used in conjunction with one another to link measurement throughout the organization thus giving visibility into the organizations progress in achieving strategic goals through process improvement [14].

The development of a core set of metrics for implementing the Balanced Scorecard is the most difficult aspect of the approach. Developing metrics that create the necessary linkages of the operational directives with the strategic mission prove to be fundamentally difficult as it is typical to view organizational performance in terms of outcomes or results rather than focus on metrics that address performance drivers that provide feedback concerning day-to-day organizational progress.

The BSC is not the organizational strategy but rather a measurement paradigm to provide operational and tactical feedback. The organizational strategic vision and goals are the foundation upon which the framework is constructed and are taken from public domain documents. The strategic plan contains the vision, goals, mission and values for the organization. The Government Performance and Results

Act, GPRA requires all federal agencies to establish strategic plans and measure their performance in achieving their missions. The vision and goals are stated below.

> Vision: To be world-class creators and facilitators of innovative, intelligent, high performance, reliable informational technologies that enable NASA missions.
>
> Goals: To become an international leading force in the field of software engineering for improving safety, reliability, quality, cost and performance of software systems; and to become a national Center of Excellence (COE) in systems and software independent verification and validation.

# 3 BSC Architecture

The BSC architecture was intended to provide a framework for industry and for-profit organizations. The framework facilitates translating the strategic plan into concrete operational terms that can be communicated throughout the organization and measured to evaluate its day-to-day viability. The three principles of building a balanced scorecard that is linked through a measurement framework to the organizational strategy include;

(1) defining the cause and effect relationships,
(2) defining the outcome measures and performance drivers,
(3) linking the scorecard to the financial outcome measures [5].

The initial steps of BSC engage in the construction of a set of hypotheses concerning cause and effect relationships among objectives for all four perspectives of the balanced scorecard. The measurement system makes these relationships explicit. Therefore, they can be used to assess and evaluate the validity of the BSC hypotheses. The questions asked in each category of the four perspectives provide a segue into the cause effect diagramming activity. It is this activity that exposes the value chain associated with specific IV&V activities.

## 3.1 Defining the Cause-Effect Relationships

IV&V is conducted using different approaches and methods depending the goals of the IV&V team. To define causal relationships we must evaluate the measurement based on a context sensitive method:

1) Identify the underlying IV&V process relative to the development process.
2) Identify the activities (methods, models and tools) by inputs and outputs and entry and exit criteria.
3) For activities categorized as information management IT, measure the value of information to decrease uncertainty, mitigate risk, improve quality...
4) For analysis activities we define the value for the outputs such as problem reports at a given time in the lifecycle and by criticality.

We begin by formulating hypotheses concerning the value of IV&V in a given context of the Space Shuttle IV&V activities. The hypotheses are based on inferred or known relationships documented in prior studies reviewed under the first phase of our ROI project. We state the initial hypotheses as constructed, however their review and evaluation are an ongoing activity.

The hypotheses developed are based on several assumptions that are based on current understanding of the interaction of the IV&V process and shuttle development process. The Space Shuttle is considered a product-line as defined by the SEI as well as the general research community. The characteristics that make the shuttle a product line process include the systematic reuse of a set of core architectural and component based assets that are reused in each incremental release. This core commonality is extended to support each operational increment (OI) and represents a negotiated and limited degree of domain variability.

> Hypothesis 1: The benefits of IV&V contributions are realized as domain engineering and applications engineering benefits. This means some benefits should accrue to the core structure of shuttle software and be an ongoing contribution in its maintenance and extensibility.
>
> Hypothesis 2: The benefits of the application engineering accrue almost entirely to the developer. That is the defect reduction that occurs in development is enabled in part by IV&V contributions to domain engineering.
>
> Hypothesis 3: The benefits of product-line engineering in the shuttle are significant in reducing testing costs while maintaining high levels of testing quality. The degree of test suite and test environment reuse is exceptionally high and results in a significant cost savings.
>
> Hypothesis 4: This is fundamentally a unique system that is developed using sophisticated reuse. This requires us to view the system as generating shuttle "builds" from an investment of core assets. The benefits are primarily derived in the reusability and rapid extensibility of the shuttle code.
>
> Hypothesis 5: Adherence to an architecture enables system safety, reliability and quality standards to be imposed and verified for the core assets of the shuttle. Acceptable degrees of variability to extend functionality are approved by a team of architects and systems engineers that includes the IV&V team.

We map our hypothesis to a set of objectives concerning the value of IV&V and the necessary and sufficient factors to creating value for the organization in terms of the strategic vision and goals. The BSC is segmented into four categories of objectives customer, financial, internal business processes and learning and growth segments. The objectives for the four segments are the following:

- customer segment objectives correspond with the high level goals of mission success through high quality, reliability and safety.
- financial segment objectives focus on cost reduction, efficient asset utilization and high ROI values of IT investments.
- internal process objectives relate to specific software and systems engineering approaches such as product-line development paradigms, CPI and QIP efforts, and test technologies and best practices as defined for IV&V.
- learning and growth objectives include technological infrastructure for distributed development, workforce training programs, skills assessment program, and ISO-9000 process structure.

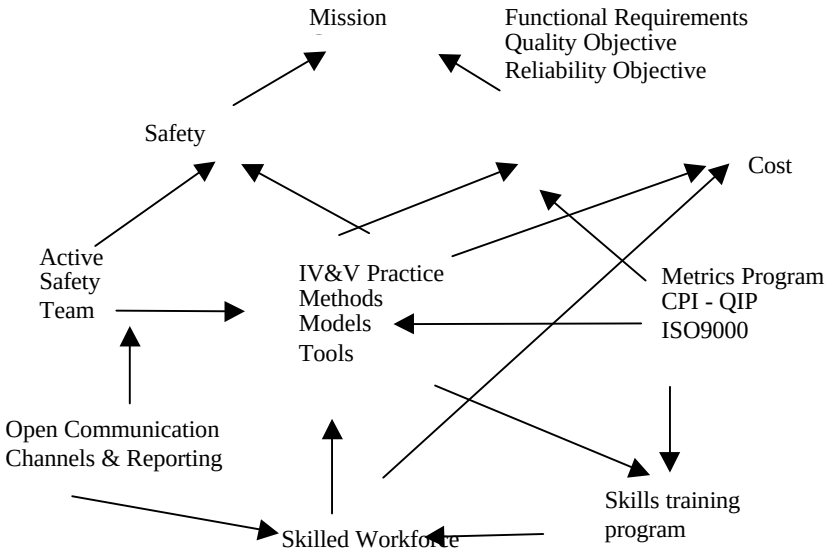


Figure 1.1 Influence diagram of IV&V BSC objectives.

The objectives are used in the selection of a minimum set of required metrics to measure day-to-day performance as well as longer term outcome or results metrics. This aspect of the framework focuses on development of leading and lagging indicators. An example customer focused objective would be the improvement in overall safety due to IV&V activities. A leading indicator for this objective could be the number of identified potential hazardous states resulting from a safety impact analysis or a tracking of the hazard rate during development. A result measure or lagging indicator could be the number of in flight anomalies (IFA) that are documented. The leading and lagging indicators must be assigned desired or normative values. These values become targets or target ranges for the metrics collected. Finally, the initiatives that have been sponsored to achieve the objective is identified and reevaluated with respect to the quantitative and qualitative evidence of success relative to the target values (see table 1.1.)

| | Objectives | Measures | Targets | Initiatives |
|---|---|---|---|---|
| Customers (Internal External) | No Losses | # Severity 1 &2 | Remove < FRR | Formal Methods |
| | Reduce Risk | # IFA's | No Severity 1 | Risk Management |
| | Manage Risk | Fault tolerance | Performance | Risk Mitigation |

Table 1.1 Customer focus metrics definition.

The relationships among the customer objectives of interest are significant as they are not independent of one another and therefore must be analyzed based on their degree of covariance and interaction. The relationships are diagrammed Fig.1.3 and depict the current accepted understanding. Safety requires that unsafe states cannot be entered from any point of function of the system. It is possible for the systems to function reliably that is without failure and still enter unsafe states of operation. A system can be completely correct and defect free and still enter unsafe states. There are many documented examples of these properties in the literature and many devoted specifically to documenting the complexity of software safety issues. The safety of a system is a result of its safe operation in a specific context or environment. We provide definitions of safety, reliability, quality and cost as defined for the customer objectives of the BSC.

- Safety is defined as freedom from accidents or losses. This is an absolute statement, safety is more practically viewed as a continuum from no accidents or losses to acceptable levels of risk of loss.
- Reliability is defined in terms of the probabilistic or statistical behavior, that is the probability that the software will operate as expected over a specified period of time.
- Quality is defined in terms of correctness and number of defects. Correctness is an absolute quality, it is also a mathematical property that establishes the equivalence between the software and its specification.
- Cost is more complex than it appears, direct or absorption costing may be applied and alters what costs are included and therefore what costs may be reduced. The focus of the paper does not rely on the differences inherent to these two approaches and therefore defers discussion of this topic.

The NASA IV&V facility must document the increase in software and systems safety, reliability and quality that are attributable to IV&V technologies. This requires that the contribution that is made towards meeting required targets through the application of IV&V activities must be quantified. This requires that each aspect be evaluated relative to some objective target. The value add of IV&V is measured as the sum of overall reduction of distance from the target. This provides a measure of overall impact to mission success. The relative reduction of “Euclidean Distance” from the safety target of no losses attributable to IV&V specifically is documented and integrated into the overall model that sums the total reduction of distance from the three targets of safety, reliability and quality. There are many measures that can be collected to evaluate the value added of IV&V for software and system safety; this is only one approach. The measurement of the contribution of IV&V in improving safety, reliability and quality while reducing cost is discussed in the following sections.

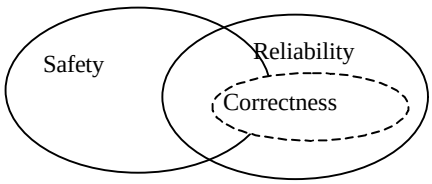


Fig.1.3 Relationships among customer themes of mission success through safety, reliability, and quality at reduced costs.

# 4 BSC Issues and Lesson Learned

The four strategic mission goals of importance to our customers are safety, reliability, quality and cost. This section discusses those aspects in terms of measurement as is defined in the balanced scorecard.

***SAFETY*** The contribution of IV&V to shuttle safety is difficult to measure directly. It is therefore necessary to make assumptions concerning those factors that would impact safety and to what degree. It is assumed that a reduction in the probability of failure is a contribution to increased safety. A reduction of the number of In Flight Anomalies IFAs of a severe nature due to IV&V identification and removal is a contribution. An independent evaluation of potential failure modes that results in identifying previously unidentified hazards is a contribution.

***RELIABILITY*** The contribution of IV&V to shuttle reliability is more directly attributable to the specific verification activities that are applied during the Shuttle software development process towards defect management. Research investigating the ramifications of testing strategies for reliability provides quantification of benefits relative to specific IV&V activities. A minimization of estimated residual faults is provided according to the sequence of testing strategies and the duration of those test executions. For example the number of defects detected by applying functional, decision, data flow and mutation test methods in sequence. The CPU execution time or the number of test cases can measure test effort. As the test effort increases defects detected can be optimized through applying more optimistic or pessimistic test strategies. The resulting increase in reliability is measured by increased MTTF or improved failure intensity profiles and is quantified as a reduction in the distance from the reliability targets of subsystems undergoing IV&V.

***QUALITY*** The contribution of IV&V to shuttle quality is measured as a reduction of defect density trends through process improvement paradigms such as traversing the CMM stages from levels 2,3,4 to level 5. The intuition behind this model is that the measurable impact of process improvement is in the reduction of the cost of rework Specific examples of applying this concept are documented in the literature and state substantial savings associated with rework avoidance. Raytheon Systems Corporation reported cost savings of $15.8 million for 15 projects over a four-year period. Raytheon documents an ROI of 7:1 based on $4.48 million return for $580,000 invested. Hughes Aircraft reported cost savings of $9.2 million over a three-year period. Hughes documents an ROI of 4.5:1 based on $2 million return on $400,000 invested. The Aircraft Software Division at Tinker Air Force Base reported an ROI of 6.35:1 based on a return of $2.9 million for $462,100 invested. In addition, the rework cost avoidance of detecting defects of severity 1; severity 2 and severity 3 can be quantified relative to phase of detection and level of severity. The reduction of defect density is measured as a reduction of distance from the overall quality objective measured in defect density according to severity.

***COST*** In the early 1990's the software engineering community adapted ROI to measure the costs and benefits of SEI/CMM process improvement efforts. Published examples of how ROI for CMM based process improvements are measured and interpreted provide guidelines for the basic proposed ROI model [7,13]. The process

community quantified process and product improvement using the following four major development-cost structures drawn from Crosby's work as published in "Quality is Free" and "Quality Without Tears" [3,4]. Crosby's work is referenced by Capers Jones as the seminal work in this area and has been used as the basis for cost structuring by DoD contractors such as Raytheon Systems [17]. The cost categories include:

1. nonconformance rework costs (such as fixing code defects or design documentation),
2. performance costs associated with doing it right the first time (such as developing the design or generating the code),
3. appraisal costs associated with testing the product to determine if its faulty, and
4. prevention costs incurred trying to prevent faults from degrading the product.

Industry has applied these four cost categories to measuring ROI for software process improvement by using rework costs avoided (nonconformance costs avoided) as the numerator and appraisal and prevention costs directly related to process improvement efforts for the denominator [7,18]. The intuition behind this model is that the measurable impact of process improvement is in the reduction of the cost of rework [3,4,10,11].

A measurement framework is necessary to bridge the gap between strategic measures of improved reliability, safety, and quality at reduced cost and operational measures of optimization of resource allocations applicable to daily activities to achieve these goals. The BSC provides a means of measuring the efficiency of resource allocations for the operational processes of software and systems verification and validation activities that must then be linked to the high level goals of mission success at reduced cost. In applying the BSC we have learned many lessons of value concerning our strategic planning as it relates to the activities conducted to accomplish daily operational goals. First, we have found that a customer focus of the strategic themes provides the necessary linkages in the BSC to measure our leading and lagging indicators successfully. We have also learned that the CMM and ISO-9000 initiatives are split across the core process tier and the infrastructure tier of the BSC hierarchy. These two findings are essential in applying the BSC to a government or not-for-profit organization such as the NASA IV&V Facility.

## 5 Future Directions

The primary focus of learning and growth measures for IV&V specifically is the information technologies (IT) used to obtain, retrieve, disseminate and store key information products [6]. The IV&V Facility is located in West Virginia and yet services all the NASA Centers from the Pacific to Atlantic coasts. To support this distributed context. Communications technologies such as VITS, VOTS and internet tools such as web-based data collection repositories are required. Specific measures to quantify performance, cost, and quality for IT infrastructure to support IV&V technologies must be further evaluated to provide meaningful target ranges for IT performance metrics.

In addition, further investigation into the measurement of core processes as defined under ISO is required. The ISO-9126 Standard, documents 6 high-level software qualities including functionality, reliability, usability, efficiency, maintainability and portability. These high-level qualities are mapped to 24 sub-characteristics. Metrics are proposed to measure the high-level software qualities relative to the sub-characteristics. This ISO standard could provide the necessary metrics to measure operational processes under the process aspect of the BSC, relative to the application of product line reuse, and map them to the high-level goals. Of particular interest in this standard is the definition of reusability as the combination of maintainability and portability. It will be of interest to analyze the appropriateness of the standard in measuring reuse for the shuttle [9]. Specifically, reuse across a vertical product line that incorporates domain engineering, architecture-based reuse, and reusable test technologies.

**REFERENCES**


[1] Basili, V. Rombach, D., "The TAME Project: Towards Improvement Oriented Software Environments," IEEE Trans. Software Engineering, 1988.

[2] Boehm, B., Software Engineering Economics, Englewood Cliffs, Prentice Hall, 1981.

[3] Crosby, P. B., Quality is Free. McGraw Hill, 1979.

[4] Crosby, P. B., Quality without Tears. McGraw Hill, 1985.

[5] Eickelmann, Nancy S., "Combining Software Measurement Frameworks to Assess the Operational and Strategic Value of Process Improvement in a Government Organization"European Software Process Improvement Conference: Learn from the past – experience the future. In the Proceedings of the EuroSPI '99 at the Pori School of Technology and Economics, Pori, Finland, Oct. 25-27, 1999.

[6] Eickelmann, Nancy S., " Strategic and Software Measurement Frameworks to Assess the Value of Information Technology", In the Proceedings of FESMA '99 European Software Measurement Conference, Amsterdam, Netherlands. Oct. 4-8, 1999.

[7] Eickelmann, Nancy S., "A Comparative Analysis of BSC as Applied in Government and Industry Organizations." Information Technology Balanced Scorecard Symposium, Antwerpen, Belgium, March 15-16, 1999.

[8] Eickelmann, Nancy S., "Measuring and Evaluating the Software Test Process." European Software Measurement Conference, FESMA `98, Antwerp, Belgium, May 6-8, 1998.

[9] Eickelmann, Nancy S., Product-Line Development Metrics. GSAW `98, El Segundo, California, February 25,1998.

Hetzel, B., Making Software Measurement Work. John Wiley and Sons, 1993.

[10] Humphrey, W., Managing the Software Process. Addison-Wesley 1989.

[11] Humphrey, W., Snyder, T., and Willis, R., "Software Process Improvement at Hughes Aircraft," IEEE Software, July 1991.

[12] Jenner, M., Software Quality Management and ISO 9000. John Wiley and Sons, 1995.

[13] Jones, C., Applied Software Measurement. McGraw Hill, 1991.

[14] Kaplan, R. and Norton, D., The Balanced Scorecard. Harvard Business School Press, 1996.

[15] McGrath, R. and MacMillan, I., "Discovery-Driven Planning" Harvard Business Review, July-August 1995.

[16] Radatz, J. W., “Analysis of IV&V Data” Rome Air Development Center ROME C# F30602-80-C-0115, 1981.
[17] Saiedian, H. and Kuzara, R., “SEI Capability Maturity Model’s Impact on Contractors” IEEE Computer, January 1995.
[18] Violino, B., “Measuring Value: Return on Investment” Information Week, Issue 637, June 30, 1997.